Support for remote and/or S3 storage

Use case: NASA Access to Terra Fusion

Remote storage

As of today, the NASA Access project has data stored primarily on Blue Waters nearline, which is accessible only via Blue Waters or Globus. Discussing possible application of Clowder in this environment, it would be helpful if it were possible to store information about the files in Clowder without necessarily needing direct access to the bytes on disk. In this sense, Clowder would function primarily as a flexible metadata repository. A researcher would then be able to process the data externally, for example via an XSEDE resource, and update metadata in Clowder about files stored on nearline. Other users would be able to discover this information via Clowder and still operated on remote resources via Globus transfer.

S3 storage

The NASA Acces project is exploring viability of processing via AWS batch. A secondary scenario would be that the basic fusion data are available via S3. Researchers would use AWS batch (or other) processing platforms to create derived data products/metadata, which could be posted back to Clowder referencing the data in S3.